

# Composite Kernel Local Angular Discriminant Analysis for Multi-Sensor Geospatial Image Analysis

Saurabh Prasad, *Senior Member, IEEE*, Minshan Cui, Lifeng Yan

**Abstract**—With the emergence of passive and active optical sensors available for geospatial imaging, information fusion across sensors is becoming ever more important. An important aspect of single (or multiple) sensor geospatial image analysis is feature extraction — the process of finding “optimal” lower dimensional subspaces that adequately characterize class-specific information for subsequent analysis tasks, such as classification, change and anomaly detection etc. In recent work, we proposed and developed an angle-based discriminant analysis approach that projected data onto subspaces with maximal “angular” separability in the input (raw) feature space and Reproducible Kernel Hilbert Space (RKHS). We also developed an angular locality preserving variant of this algorithm. In this letter, we advance this work and make it suitable for information fusion — we propose and validate a composite kernel local angular discriminant analysis projection, that can operate on an ensemble of feature sources (e.g. from different sources), and project the data onto a unified space through composite kernels where the data are maximally separated in an angular sense. We validate this method with the multi-sensor University of Houston hyperspectral and LiDAR dataset, and demonstrate that the proposed method significantly outperforms other composite kernel approaches to sensor (information) fusion.

**Index Terms**—hyperspectral data, dimensionality reduction, composite kernels pattern recognition

## I. INTRODUCTION

Optical remote sensing has made significant advances in recent years. Among these has been the deployment and widespread use of hyperspectral imagery on a variety of platforms (including manned and unmanned aircraft and satellites) for a wide variety of applications, ranging from environmental monitoring, ecological forecasting, disaster relief to applications pertaining to national security. With rapid advancements in sensor technology, and the resulting reduction of size, weight and power requirements of the imagers, it is also now common to deploy multiple sensors on the same platform for multi-sensor imaging. As a specific example, it is appealing for a variety of remote sensing applications to acquire hyperspectral imagery and Light Detection and Ranging (LiDAR) data simultaneously — hyperspectral imagery offers a rich characterization of object specific properties, while LiDAR provides topographic information that complements Hyperspectral imagery [1]–[5]. Modern LiDAR systems provide the ability to record entire waveforms for every return signal as

opposed to providing just the point cloud. This enables a richer representation of surface topography.

While feature reduction is an important preprocessing to analysis of single-sensor high dimensional passive optical imagery (particularly hyperspectral imagery), it becomes particularly important with multi-sensor data where each sensor contributes to high dimensional raw features. A variety of feature projection approaches have been used for feature reduction, including classical approaches such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) and their many variants, manifold learning approaches such as Supervised and Unsupervised Locality Preserving Projections [6]. Several of these methods are implemented in both the input (raw) feature space and the Reproducible Kernel Hilbert Space (RKHS) for data that are nonlinearly separable. Further, most traditional approaches to feature extraction are designed for single-sensor data — a unique problem with multi-sensor data is that feature spaces corresponding to each sensor often have different statistical properties, and a single feature projection may hence be sub-optimal. It is hence desired to have a projection for feature reduction that preserves the underlying information from each sensor in a lower dimensional subspace.

More recently, we developed a feature projection approach, referred to as Angular Discriminant Analysis (ADA) [7]–[9], that was optimized for hyperspectral imagery and demonstrated robustness to spectral variability. Specifically, the approach sought a lower dimensional subspace where classes were maximally separated in an angular sense, preserving important spectral shape related characteristics. We also developed a local variant of the algorithm (LADA) that preserved angular locality in the subspace. In this paper, we propose a composite kernel implementation of this framework and demonstrate for the purpose of feature projection in multi-sensor frameworks. Specifically, by utilizing a composite kernel (a dedicated kernel for each sensor), and ADA (or LADA) for each sensor, the resulting projection is highly suitable for classification. The proposed approach serves as a very effective feature reduction algorithm for sensor fusion — it optimally fuses multi-sensor data and projects it to a lower dimensional subspace. A traditional classifier can be employed following this, for supervised learning. We validate the method with the University of Houston multi-sensor dataset comprising of Hyperspectral and LiDAR data and show that the proposed method significantly outperforms other approaches to feature fusion.

The authors are with the Department of Electrical & Computer Engineering at the University of Houston. This material is based upon work funded in part by the NASA New Investigator (Career) Award, Grant No. NNX14A147G. Corresponding author: Saurabh Prasad (email: saurabh.prasad@ieee.org).

The outline of the remainder of this paper is as follows. In sec. II, we review related work. In sec. III, we describe the proposed approach for multi-sensor feature extraction. In sec. IV, we describe the experimental setup and present results with the proposed method, comparing it to several state-of-the-art techniques to feature fusion.

## II. RELATED WORK

Traditional approaches to feature projection based dimensionality reduction such as PCA, LDA and their variants largely rely on Euclidean measures. Manifold learning approaches [6] also seek to preserve manifold structures and neighborhood locality through projections that preserve such structures. Other projection based approaches to feature reduction, such as Locality Preserving Projections (LPP), Local Fisher's Discriminant Analysis (LFDA) [6], [10], [11] etc. integrate ideas of local neighborhoods through affinity matrices, into classical projection based analysis approaches such as PCA, LDA etc. As a general feature extraction approach, Euclidean distance is a reasonable choice, including for remotely sensed image analysis. However, by noting the well understood benefits of spectral angle for hyperspectral image analysis, in previous work, we developed an alternate feature projection paradigm that worked with angular distance measures instead of euclidean distance measures [7]–[9] — we demonstrated that when projecting hyperspectral data through this class of transformations, the resulting subspaces were very effective for downstream classification and significantly outperformed their Euclidean distance counterparts. In addition to benefits with classification, we also demonstrated other benefits of this class of methods, including robustness to illumination differences — something that is very important for remote sensing. In other previous work, it has been shown that a reproducible kernel Hilbert space (RKHS) generated by composite kernels (a weighted linear combination of basis kernels) is very effective for multi-source fusion [2], [12]–[14]. Here, we briefly review the developments related to angular discriminant analysis. This will provide a context and motivation for the proposed work in this paper that seeks to demonstrate the benefits of composite kernel angular discriminant analysis for multi-source image analysis.

### A. Angular Discriminant Analysis

Here, we briefly review Angular Discriminant Analysis (ADA) and its locality preserving counterpart, Local Angular Discriminant Analysis (LADA). Consider a  $d$ -dimensional feature space (e.g. hyperspectral imagery with  $d$  spectral channels). Let  $\{x_i \in \mathbb{R}^d, y_i \in \{1, 2, \dots, c\}\}$  be the  $i$ -th training sample with an associated class label  $y_i$ , where  $c$  is the number of classes. The total number of training samples in the library is  $n = \sum_{l=1}^c n_l$ , where  $n_l$  denotes the number of training samples from class  $l$ . Let  $T \in \mathbb{R}^{d \times r}$  be the desired projection matrix, where  $r$  denotes the reduced dimensionality. We also denote symbols having unit norm with a *tilde* — this will be useful where we normalize the data to a unit norm to focus on angular separability.

1) *ADA*: Traditional LDA seeks to find a subspace that maximizes between-class scatter while minimizing within-class scatter, where the scatter is measured using Euclidean distances. While similar in philosophy, ADA is an entirely new approach to subspace learning that is based on *angular* scatter — it seeks a subspace where within-class angular scatter is maximized, and the between-class angular scatter is maximized. Just like LDA, the ADA optimization problem can be posed as a generalized eigenvalue problem. Specifically, ADA seeks to find a projection where the ratio of between-class inner product to within-class inner product of data samples is minimized. The within-class outer product matrix  $O^{(w)}$  and between-class outer product matrix  $O^{(b)}$  are defined as

$$O^{(w)} = \sum_{l=1}^c \sum_{i: y_i=l} \tilde{\mu}_l \tilde{x}_i^t, \quad (1)$$

$$O^{(b)} = \sum_{l=1}^c n_l \tilde{\mu} \tilde{\mu}_l^t. \quad (2)$$

where  $\tilde{\mu}_l = \frac{1}{n_l} \sum_{i: y_i=l} \tilde{x}_i$  is the normalized mean of  $l$ -th class samples, and  $\tilde{\mu} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i$  is defined as the normalized total mean.

It was shown in [7] that the projection matrix  $T_{ADA}$  of ADA can be approximated as the solution to the following trace ratio problem

$$T_{ADA} \approx \underset{T \in \mathbb{R}^{d \times r}}{\operatorname{argmin}} \left[ \operatorname{tr} \left( (T^t O^{(w)} T)^{-1} T^t O^{(b)} T \right) \right]. \quad (3)$$

The projection matrix  $T_{ADA}$  can be obtained by solving the generalized eigenvalue problem involving  $O^{(w)}$  and  $O^{(b)}$ .

2) *LADA*: Similar to LDA, ADA is a “global” projection in that it does not specifically promote preservation of local (neighborhood) angular relationships under the projection. We hence developed LADA in [7], which is a local variant of ADA. The within and between-class outer product matrices of LADA are obtained as follows

$$O^{(lw)} = \sum_{i,j=1}^n \tilde{W}_{ij}^{(lw)} \tilde{x}_i \tilde{x}_j^t, \quad (4)$$

$$O^{(lb)} = \sum_{i,j=1}^n \tilde{W}_{ij}^{(lb)} \tilde{x}_i \tilde{x}_j^t, \quad (5)$$

where the normalized weight matrices are defined as

$$\tilde{W}_{ij}^{(lw)} = \begin{cases} \tilde{A}_{ij}/n_l, & \text{if } y_i, y_j = l, \\ 0, & \text{if } y_i \neq y_j, \end{cases} \quad (6)$$

$$\tilde{W}_{ij}^{(lb)} = \begin{cases} \tilde{A}_{ij}(1/n - 1/n_l), & \text{if } y_i, y_j = l, \\ 1/n, & \text{if } y_i \neq y_j. \end{cases} \quad (7)$$

The normalized affinity matrix  $\tilde{A}_{ij} \in [0, 1]$  between  $\tilde{x}_i$  and  $\tilde{x}_j$  is defined as

$$\tilde{A}_{ij} = \exp \left( - \frac{(2 - 2\tilde{x}_i^t \tilde{x}_j)}{\tilde{\gamma}_i \tilde{\gamma}_j} \right), \quad (8)$$

where  $\tilde{\gamma}_i = \sqrt{2 - 2\tilde{x}_i^t \tilde{x}_i^{(knn)}}$  denotes the *local angular scaling* of data samples in the angular neighborhood of  $\tilde{x}_i$ , and  $\tilde{x}_i^{(knn)}$  is the  $K$ -th nearest neighbors of  $\tilde{x}_i$ .

Similar to ADA, the projection matrix of LADA can be defined as

$$T_{LADA} = \operatorname{argmin}_{T \in \mathbb{R}^{d \times r}} \left[ \operatorname{tr} \left( (T^t O^{(lw)} T)^{-1} T^t O^{(lb)} T \right) \right]. \quad (9)$$

### III. COMPOSITE KERNEL ANGULAR DISCRIMINANT ANALYSIS FOR IMAGE FUSION

In this section, we develop and describe the proposed approach to multi-source feature extraction — composite kernel angular discriminant analysis (CKADA) and its locality preserving counterpart (CKLADA). Our underlying hypothesis with this work is that even when angular information is important for optical image analysis, in a multi-source (e.g. multi-sensor scenario), having dedicated kernels (specific to each source) would result in a superior projection that addresses source-specific nonlinearities. With that goal, we extend our previous work with angular discriminant analysis by implementing it in a composite kernel reproducible kernel Hilbert space and demonstrate for a specific application of multi-sensor image analysis that the resulting subspace is highly discriminative and outperforms other subspace learning approaches.

Consider a nonlinear mapping  $\Phi(\cdot)$  from the input space to a RKHS  $\mathcal{H}$  as follows:

$$\Phi : \mathbb{R}^d \rightarrow \mathcal{H}, x \rightarrow \Phi(x). \quad (10)$$

and a kernel function  $K$  defined as:

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle, \quad (11)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product of two vectors. Consider next a set of  $M$  co-registered multi-source images resulting in the following  $M$ -Tuple of feature vectors from co-registered images for every geolocation (co-registered pixels):  $\{x^1, x^2, \dots, x^M\}$ , where  $x^m \in \mathbb{R}^{d_m}$ . Associated with every pixel (geolocation) for which ground truth is available, there is a class label  $y \in \{1, 2, \dots, c\}$ . A composite kernel RKHS can then be constructed as

$$K(x_i, x_j) = \sum_{m=1}^M \alpha_m K_m(x_i^m, x_j^m) \quad (12)$$

where  $K_m$  is a basis kernel for the  $m$ 'th source, formed by any valid Mercer's kernel. To implement Composite Kernel ADA (CKADA), note that  $O^{(w)}$  and  $O^{(b)}$  can be reformulated as

$$O^{(w)} = XW^{(w)}X^t, \quad (13)$$

$$O^{(b)} = XW^{(b)}X^t. \quad (14)$$

where  $W^{(w)}$  is given as

$$W_{ij}^{(w)} = \begin{cases} 1/n_l, & \text{if } y_i, y_j = l, \\ 0, & \text{if } y_i \neq y_j. \end{cases} \quad (15)$$

and  $W^{(b)}$  is given as

$$W_{ij}^{(b)} = \begin{cases} 1/n - 1/n_l, & \text{if } y_i, y_j = l, \\ 1/n, & \text{if } y_i \neq y_j. \end{cases} \quad (16)$$

ADA can hence be re-expressed as the solution to the following generalized eigenvalue problem

$$XW^{(b)}X^t\nu = \lambda XW^{(w)}X^t\nu. \quad (17)$$

Since  $\nu$  can be represented as a linear combination of columns of  $X$ , it can be formulated using a vector  $\varphi \in \mathbb{R}^n$  as

$$X^t\nu = X^tX\varphi = K\varphi, \quad (18)$$

where  $K$  is a  $n \times n$  symmetric kernel (Gram) matrix. Here  $K_{ij} = \kappa(x_i, x_j) = \langle x_i, x_j \rangle$  represents a simple inner product kernel, but can be replaced by (12) by utilizing the kernel trick. Multiplying  $X^t$  on both sides of (17), results in the following generalized eigenvalue problem.

$$KW^{(b)}K\varphi = \lambda KW^{(w)}K\varphi. \quad (19)$$

Let  $\Psi = \{\varphi_k\}_{k=1}^r$  be the  $r$  generalized eigenvectors associated with the  $r$  smallest eigenvalues  $\lambda_1 \leq \lambda_2, \dots, \leq \lambda_r$ . A test sample  $x_{test}$  can be embedded in  $\mathcal{H}$  via

$$(X\Psi)^t x_{test} = \Psi^t X^t x_{test} = \Psi^t K_{X, x_{test}}, \quad (20)$$

where  $K_{X, x_{test}}$  is a  $n \times 1$  vector. Composite Kernel Local ADA (CKLADA) can likewise be implemented by replacing the weight matrices ( $W^{(w)}$  and  $W^{(b)}$ ) above with their local counterparts defined in (6) and (7).

We note that in the proposed approach, the empirical kernel (Gram) matrix from (12) that is formed as a weighted linear combination over all sources is used in the generalized eigenvalue problem for CKLADA (20). The algorithm projects the data from  $M$  sources onto a *unified RKHS* through a bank of kernels individually optimized for each source. The final embedding seeks to optimally separate (in an angular sense) data in the RKHS. The linear mixture of kernel enables us to optimize each kernel (for example the kernel parameters) for each source instead of applying a single kernel for all sources, and to specify source importance (via mixing weights) to the overall analysis task at hand.

*Practical Considerations:* We note the following free parameters in the overall embedding that affect the subspace that is generated: Embedding dimension,  $r$ , mixture weights used in the composite kernel,  $\{\alpha_m\}_{m=1}^M$ , choice of kernel and related kernel parameters. We note that unlike some other embedding techniques such as LDA and its variants where the embedding dimension is upper bounded due to rank deficiency of the between class scatter, with composite kernel local ADA, the between class angular scatter is not rank limited, and as a result, the projection matrix resulting from the solution to the generalized eigenvalue problem does not enforce an upper bound on the embedding dimension. Hence,  $r$  is a free parameter that represents the *unified subspace* generated by all  $M$  sources. The choice of  $r$  should hence be governed by the information content (as quantified for example in the eigenspectra of the decomposition). The choice of weights can be made through cross validation or techniques such as kernel alignment — in our experience, there is often a very wide plateau over a range of values of the weights, and hence we chose to use simple cross validation to learn weights from our training data. We utilized a standard radial basis

function (RBF) kernel for each source ( $K_m$ ), but the kernel parameter (width of the RBF kernel) is optimized for each source individually via cross validation.

*Classification:* We note that following a CKLADA projection, a simple classifier can be utilized for down-stream analysis. This follows from the observation that applying kernel projections while simultaneously ensuring preservation of angular locality will result in subspaces where class-specific data are compactly clustered. We validate and measure the efficacy of subspaces resulting from CKLADA by utilizing the following classifiers: (1) A K Nearest neighbor (KNN) classifier, (2) A Gaussian maximum-likelihood (ML) classifier, and a (3) sparse representation based classifier (SRC) [15]. While the choice of KNN and ML are obvious for subspaces formed by Kernel projections as noted in [2], we make a remark on choice of SRC as an additional classifier to measure efficacy of subspaces — this choice is motivated not only by the observation that SRC has emerged as a powerful classification approach for high dimensional remote sensing data [16]–[20] and that it exploits the inherent sparsity when representing samples using training dictionaries, but also because popular solvers used (e.g. Orthogonal Matching Pursuit, OMP) are driven by inner products to learn the sparse representation and hence they essentially exploit angular information. They implicitly seek a representation where a test sample is represented sparsely in a dictionary of training data such that the atoms that eventually contribute (have non-zero, significant representation coefficients) to the representation are angularly similar to the test data samples. We hence contend that CKLADA is particularly well suited for SRC and its variants.

#### IV. EXPERIMENTAL SETUP AND RESULTS

##### A. Dataset

The dataset we utilize represents a sensor fusion scenario, comprising of LiDAR pseudo-waveforms, and a hyperspectral image cube, and is popular in the remote sensing community as a benchmark. The data were acquired over the University of Houston campus and the neighboring urban area. All the images are at the same spatial resolution of 2.5 m and have the same spatial size of  $349 \times 1340$ . The hyperspectral image was acquired with the ITRES CASI sensor, containing 144 spectral bands, ranging from 380 nm to 1050 nm. The LiDAR DSM data was acquired by an Optech Gemini sensor and then co-registered to the hyperspectral image. The laser pulse wavelength and repetition rate were 1064 nm and 167 kHz, respectively. The instrument can make up to 4 range measurements. The total number of ground reference samples is 2832, covering 15 classes of interest, with approximately 200 samples for each class — these were determined by photo-interpretation of high resolution optical imagery.. The groundtruth map is overlaid with the gray scale image showing one channel of the hyperspectral image in Fig. 1. From the dense LiDAR point cloud, a pseudo-waveform was generated for each geolocation, that is co-registered with the hyperspectral image. The pseudo-waveform was generated by quantizing the elevation into uniform sized bins, and determining the

average intensity of points as a function of elevation bins. This provides us with a co-registered cube of waveform-like LiDAR data that is coregistered with our hyperspectral image. We note that like spectral reflectance profiles that have unique shapes depending on the material in the pixel, shapes of pseudo-waveform also correlate with the material and topographic properties in the image. Hence, angular measures (such as provided by CKLADA) would be appropriate for such analysis compared to Euclidean measures.

##### B. Baselines

To validate the efficacy of the subspaces generated by CKLADA and CKADA, we setup classification experiments using the University of Houston multi-sensor dataset. We used popular and commonly employed embeddings as baselines to compare against, including CKLFDA and KPCA. Each of these embeddings was used with 3 classifiers: KNN, ML and SRC. We note that CKLFDA is a composite kernel counterpart of LFDA based on Euclidean distance measures, and is the best possible multi-source embedding that can be compared with CKLADA — a comparison of CKLFDA vs. CKLADA provides a direct understanding of the benefits of angular information for multi-source embeddings, and of the resulting algorithmic framework proposed in sec. III.

With CKLFDA, CKADA and CKLADA, we treat each sensor (hyperspectral imagery and pseudo-waveform LiDAR) as a source, each getting its dedicated base kernel. With a single kernel KPCA, we stack features from the two sensors and project them via a single transformation based on these methods. In all cases, we use RBF kernels as our base kernels, and the width of the kernel is determined via cross validation. Other free parameters including sparsity level used in SRC, number of nearest neighbors in  $K - NN$  are also determined empirically from the training data via cross-validation.

##### C. Visualization of Embeddings

To provide a visual demonstration on the power of composite kernel angular discriminant analysis for geospatial image analysis, we provide visualization of composite kernel projections CKLADA, CKADA (both angular discriminant analysis) and CKLFDA. These results are depicted in fig. 2. The figure depicts false color images generated by projecting the multi-sensor data onto the first three most significant eigenvectors learned from CKLADA, CKADA and CKLFDA respectively. It can be clearly seen that CKLADA (and CKADA to some degree) preserve object specific properties throughout the image (for example, the highly textured objects such as urban vegetation, residential areas etc. have their spatial context significantly preserved in the lower dimensional subspace). On the contrary, CKLFDA, which can be considered as the closest benchmark/baseline competitor does not perform as well. Further, towards the right corner of the image, we point the reader to the substantial benefit of CKLADA under cloud shadows - spatial structures under cloud shadows are visible under CKLADA (and CKADA to some extent), but not when using CKLFDA.



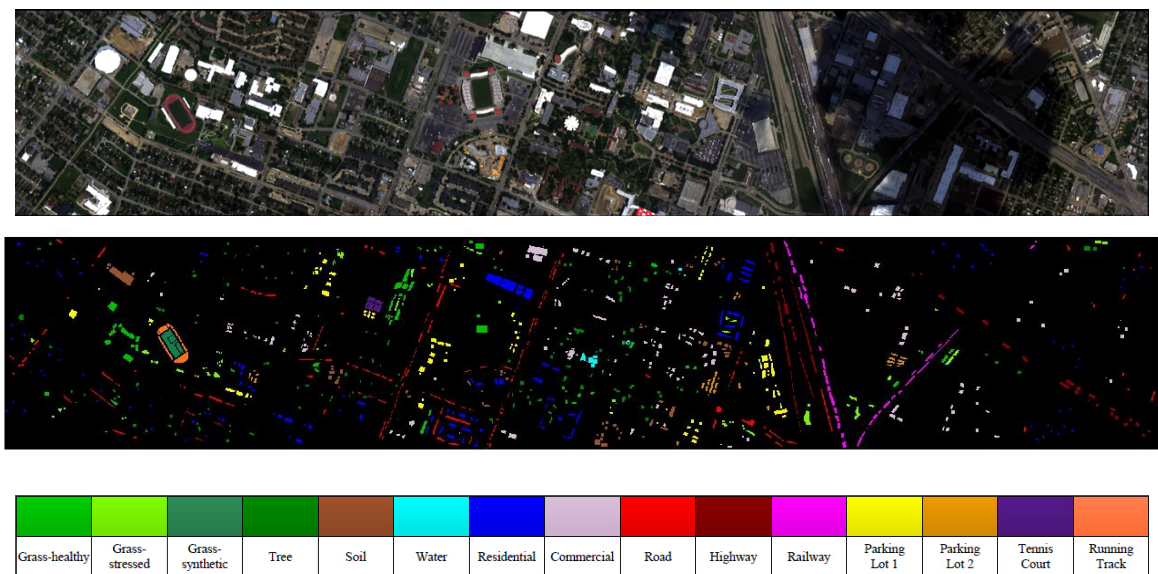
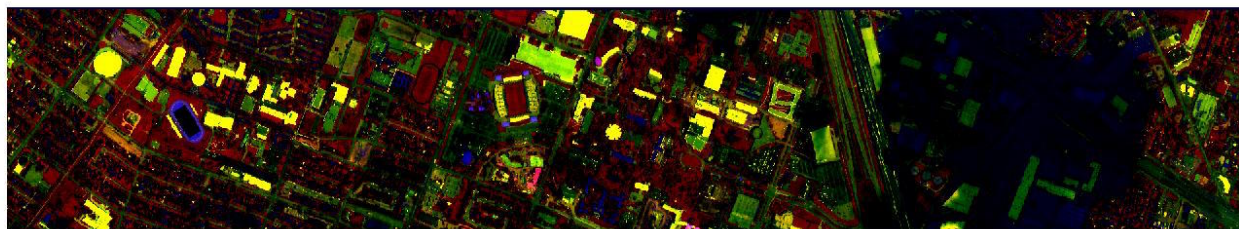


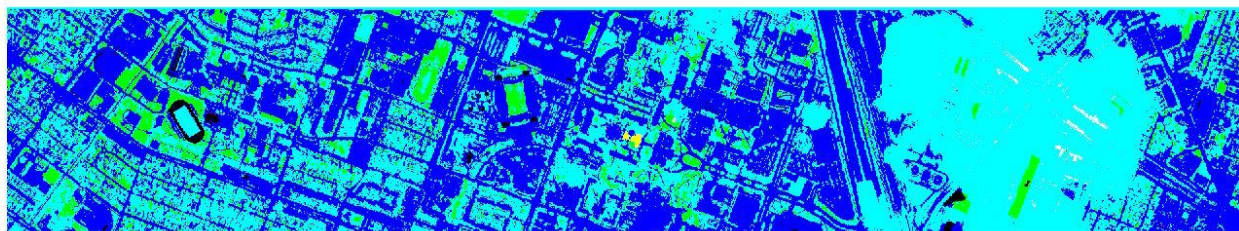
Fig. 1: True color image of hyperspectral University of Houston data, and the ground truth.



(a) CKLADA (Proposed)



(b) CKADA (Proposed)



(c) CKLFDA [2]

Fig. 2: Visualizing the projections as false color images resulting from projecting the multi-sensor data onto the first 3 significant eigenvectors using (a) CKLADA (proposed); (b) CKADA (proposed); and (c) CKLFDA.

#### D. Comparative Results

Experimental results comparing performance of CKADA and CKLADA with baseline embeddings are provided. As mentioned previously, free parameters were determined empirically via cross-validation. Tab. I depicts overall accuracy as a function of training sample size, ranging from a small to a sufficiently large value for the proposed and baseline embeddings with various classifiers. We notice that the proposed composite kernel angular discriminant analysis approaches (CKADA and CKLADA) provide anywhere from 5 – 10% improvement in performance compared to state of the art (CKLFDA), and provide even higher accuracies compared to a traditional single-kernel baseline, KPCA. We note that even when using very limited training data, CKLADA is able to substantially outperform other composite kernel and single-kernel methods (using just 10 samples per class, we obtain as much as a 10% gain in performance with CKLADA). Even when using a sufficiently large training sample set (e.g. 50 samples per class), CKLADA and CKADA outperform other methods. In Tab. II, Tab. III, and Tab. IV, we depict class specific accuracies, overall and average accuracies using the proposed methods and baselines using SRC, ML and KNN classifiers respectively. Once again, it is clear that CKADA and CKLADA consistently provide robust classification, particularly for the “difficult” classes (such as residential buildings, commercial buildings, roads, parking lots etc.). We also provide classification maps in fig. 3 using the proposed method (CKLADA) and its closest competitor, CKLFDA, using the SRC classifier. We note that CKLADA results in a map with very little noise and misclassifications, and is particularly robust under the very challenging area in the right corner of the image that is under a cloud shadow (for e.g., when using CKLFDA, the area under a cloud shadow get systematically misclassified as water — something that is visibly remedied by CKLADA). The improvements to misclassifications occurring over difficult classes is even more apparent with these ground cover classification maps.

#### V. CONCLUSIONS

We presented a composite kernel variant of angular discriminant analysis and local angular discriminant analysis. Angular discriminant analysis was previously shown to be very beneficial for high dimensional hyperspectral classification. In this paper, we expanded those developments via a composite kernel and demonstrated that this paradigm can be a very useful feature embedding algorithm in multi-source scenarios, such as when fusing multiple geospatial images. We validated our results with a popular multi-sensor benchmark and demonstrated that composite kernel angular discriminant analysis consistently outperforms other feature embeddings.

#### REFERENCES

- [1] M. Dalponte, L. Bruzzone, and D. Gianelle, “Fusion of hyperspectral and lidar remote sensing data for classification of complex forest areas,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1416–1427, 2008.
- [2] Y. Zhang and S. Prasad, “Locality preserving composite kernel feature extraction for multi-source geospatial image analysis,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 3, pp. 1385–1392, March 2015.
- [3] R. Brennan and T. Webster, “Object-oriented land cover classification of lidar-derived surfaces,” *Canadian Journal of Remote Sensing*, vol. 32, no. 2, pp. 162–172, 2006.
- [4] M. Shimoni, G. Tolt, C. Perneel, and J. Ahlberg, “Detection of vehicles in shadow areas using combined hyperspectral and lidar data,” in *Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2011, pp. 4427–4430.
- [5] M. Pedernana, P. R. Marpu, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, “Fusion of hyperspectral and lidar data using morphological attribute profiles,” in *SPIE Remote Sensing*. International Society for Optics and Photonics, 2011, pp. 81 801G–81 801G.
- [6] D. Lunga, S. Prasad, M. Crawford, and O. Ersoy, “Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning,” *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 55–66, 2014.
- [7] M. Cui and S. Prasad, “Angular discriminant analysis for hyperspectral image classification,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 9, no. 6, pp. 1003–1015, 2015.
- [8] S. Prasad and M. Cui, “Sparse representations for classification of high dimensional multi-sensor geospatial data,” in *Proceedings of the 2013 Asilomar Conference on Signals, Systems and Computers*, November 2013, pp. 811–815.
- [9] M. Cui and S. Prasad, “Sparsity promoting dimensionality reduction for classification of high dimensional hyperspectral images,” in *38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013, pp. 2154–2158.
- [10] S. Prasad, M. Cui, W. Li, and J. Fowler, “Segmented mixture of gaussian classification for robust sub-pixel hyperspectral ATR,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 138–142, January 2014.
- [11] M. Cui, S. Prasad, W. Li, and L. Bruce, “Locality preserving genetic algorithms for spatial-spectral hyperspectral image classification,” *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 6, no. 3, pp. 1688–1697, 2013.
- [12] Z. Wang and X. Sun, “Multiple kernel local fisher discriminant analysis for face recognition,” *Signal Processing*, vol. 93, no. 6, pp. 1496–1509, 2013.
- [13] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, “Multisource composite kernels for urban-image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 88–92, 2010.
- [14] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, “Composite kernels for hyperspectral image classification,” *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.
- [15] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 2, pp. 210–227, 2009.
- [16] M. Cui and S. Prasad, “Class dependent sparse representation classifier for robust hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2683–2695, May 2015.
- [17] X.-T. Yuan, X. Liu, and S. Yan, “Visual classification with multitask joint sparse representation,” *Image Processing, IEEE Transactions on*, vol. 21, no. 10, pp. 4349–4360, 2012.
- [18] Y. Chen, N. M. Nasrabadi, and T. D. Tran, “Hyperspectral image classification using dictionary-based sparse representation,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 10, pp. 3973–3985, 2011.
- [19] Y. Chen, N. Nasrabadi, and T. Tran, “Sparse representation for target detection in hyperspectral imagery,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 3, pp. 629–640, 2011.
- [20] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 471–478.

TABLE I: Comparison of various feature embedding algorithms for multi-sensor image analysis as a function of training sample size

Method	Number of training samples per class				
	10	20	30	40	50
CKLADA-SRC	73.2±3.8	87.1±1.1	90.4±1.1	92.3±0.9	93.3±0.9
CKADA-SRC	74.5±2.3	85.8±1.2	89.2±1.2	91.2±1.1	92.3±0.8
CKLFDA-SRC	66.6±2.5	81.6±1.6	86.2±1.3	87.8±1.0	88.9±1.0
KPCA-SRC	67.9±1.4	79.3±1.3	84.1±1.1	86.7±0.8	88.5±0.9
CKLADA-ML	74.33±2.4	85.7±1.6	91.1±1.2	93.3±1.0	94.3±0.7
CKADA-ML	77.2±2.7	86.7±1.6	91.6±1.2	93.0±1.0	94.0±0.8
CKLFDA-ML	70.4±2.4	81.2±1.7	86.7±1.6	88.9±1.1	90.2±1.0
KPCA-ML	72.15±2.7	85.1±1.9	91.2±1.2	93.3±1.0	94.3±0.8
CKLADA-KNN	80.3±1.7	88.1±1.3	91.4±1.0	93.0±0.9	93.9±0.7
CKADA-KNN	80.3±1.6	87.0±1.2	90.1±1.0	91.4±0.9	92.4±0.8
CKLFDA-KNN	70.7±2.2	82.5±1.6	86.5±1.2	88.1±1.0	89.3±0.9
KPCA-KNN	69.7±1.6	79.2±1.1	83.7±1.3	86.4±0.9	87.8±0.9

TABLE II: Using proposed and baseline feature embeddings with SRC

Class	Samples		Methods			
	Train	Test	CKLADA	CKADA	CKLFDA	KPCA
Grass-healthy	30	168	99.4±1.0	99.0±2.0	98.6±1.7	98.0±1.8
Grass-stressed	30	160	97.7±1.4	96.0±1.6	98.0±1.2	95.9±3.1
Grass-synthetic	30	162	99.7±0.5	99.7±0.4	95.8±2.5	98.3±1.3
Tree	30	158	98.1±1.2	95.9±2.7	98.8±1.2	99.5±0.5
Soil	30	156	99.8±0.3	98.6±0.9	98.5±1.7	97.3±2.4
Water	30	152	98.7±2.3	95.1±3.0	97.0±2.3	96.9±2.3
Residential	30	166	85.8±5.5	81.3±5.0	80.3±4.6	77.1±6.1
Commercial	30	161	86.8±5.5	82.9±6.4	77.2±6.4	79.5±5.6
Road	30	163	80.2±5.7	78.5±6.1	73.7±5.5	64.4±5.1
Highway	30	161	90.9±3.2	90.8±3.7	86.9±3.9	77.9±5.2
Railway	30	151	88.5±3.9	86.2±4.5	82.6±4.1	76.1±5.8
Parking Lot 1	30	162	75.2±5.7	74.2±6.3	67.8±6.4	62.3±5.0
Parking Lot 2	30	154	65.5±4.7	74.5±4.5	58.4±6.8	50.7±5.7
Tennis Court	30	151	99.5±0.5	98.5±1.1	98.6±2.0	96.6±2.2
Running Track	30	157	98.9±0.5	98.7±0.7	98.4±0.9	99.8±0.4
OA	–	–	91.0±1.0	89.9±0.9	87.3±1.2	84.7±1.1
AA	–	–	91.0±2.8	90.0±3.2	87.4±3.4	84.7±3.5

TABLE III: Using proposed and baseline feature embeddings with Gaussian ML

Class	Samples		Methods			
	Train	Test	CKLADA	CKADA	CKLFDA	KPCA
Grass-healthy	30	168	94.9±5.4	95.1±4.1	92.6±5.2	95.9±4.4
Grass-stressed	30	160	99.4±0.8	97.8±1.9	99.8±0.3	98.6±1.9
Grass-synthetic	30	162	96.5±0.3	96.8±2.7	88.2±4.8	96.7±2.7
Tree	30	158	99.7±0.8	99.5±0.8	98.9±1.7	99.9±0.3
Soil	30	156	97.6±2.7	93.2±4.4	95.0±4.4	97.4±2.5
Water	30	152	95.1±2.9	92.6±4.0	94.0±3.5	96.0±2.8
Residential	30	166	86.1±6.9	80.8±6.9	82.4±7.9	82.2±7.8
Commercial	30	161	86.1±8.3	91.3±6.6	73.5±11.8	86.5±9.7
Road	30	163	83.5±8.6	87.8±6.8	76.8±7.2	82.6±8.9
Highway	30	161	82.8±8.3	83.9±6.9	76.7±7.8	83.4±8.3
Railway	30	151	84.5±7.7	81.8±7.6	80.6±8.0	83.6±8.6
Parking Lot 1	30	162	71.7±8.7	49.8±11.6	64.5±7.6	71.8±10.2
Parking Lot 2	30	154	92.2±3.3	96.7±1.5	84.8±6.7	92.5±4.3
Tennis Court	30	151	98.0±2.0	96.7±3.4	95.5±2.8	98.5±1.9
Running Track	30	157	96.6±2.0	97.1±1.9	94.9±2.9	97.9±1.5
OA	–	–	90.9±1.2	89.3±1.3	86.5±1.3	84.6±1.6
AA	–	–	91.0±4.7	89.4±4.7	86.6±5.5	90.9±5.0

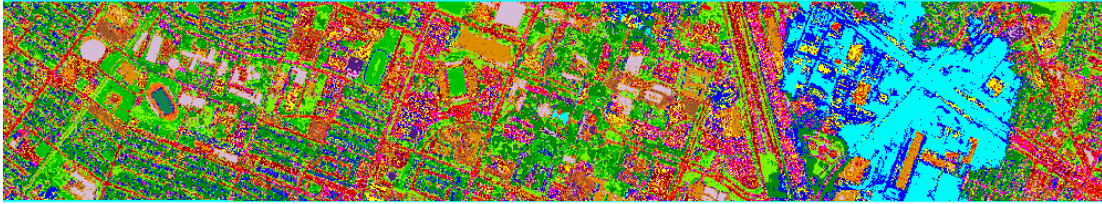


TABLE IV: Using proposed and baseline feature embeddings with KNN

Class	Samples		Methods			
	Train	Test	CKLADA	CKADA	CKLFDA	KPCA
Grass-healthy	30	168	98.6 $\pm$ 3.0	99.4 $\pm$ 2.3	97.4 $\pm$ 2.9	98.2 $\pm$ 2.8
Grass-stressed	30	160	97.5 $\pm$ 1.5	96.9 $\pm$ 1.4	97.3 $\pm$ 2.0	97.2 $\pm$ 2.6
Grass-synthetic	30	162	99.5 $\pm$ 0.8	99.7 $\pm$ 0.4	96.9 $\pm$ 1.9	97.9 $\pm$ 1.6
Tree	30	158	98.1 $\pm$ 1.8	96.0 $\pm$ 2.7	96.9 $\pm$ 2.9	99.7 $\pm$ 0.4
Soil	30	156	99.7 $\pm$ 0.5	98.5 $\pm$ 0.8	98.1 $\pm$ 2.6	98.1 $\pm$ 1.1
Water	30	152	98.4 $\pm$ 2.4	94.8 $\pm$ 2.4	97.3 $\pm$ 2.2	96.4 $\pm$ 1.7
Residential	30	166	84.6 $\pm$ 5.2	79.6 $\pm$ 6.3	73.8 $\pm$ 6.0	64.4 $\pm$ 6.7
Commercial	30	161	82.1 $\pm$ 8.1	76.4 $\pm$ 8.2	77.5 $\pm$ 7.8	79.2 $\pm$ 6.7
Road	30	163	83.1 $\pm$ 4.9	78.7 $\pm$ 5.1	74.0 $\pm$ 5.8	64.3 $\pm$ 5.3
Highway	30	161	94.3 $\pm$ 2.7	93.3 $\pm$ 3.2	89.0 $\pm$ 3.4	80.8 $\pm$ 3.8
Railway	30	151	93.4 $\pm$ 3.2	90.3 $\pm$ 4.2	83.9 $\pm$ 5.3	76.0 $\pm$ 6.3
Parking Lot 1	30	162	78.8 $\pm$ 5.9	76.6 $\pm$ 6.4	70.7 $\pm$ 6.5	67.8 $\pm$ 6.5
Parking Lot 2	30	154	65.4 $\pm$ 5.0	69.6 $\pm$ 4.1	54.2 $\pm$ 4.9	41.0 $\pm$ 4.7
Tennis Court	30	151	99.6 $\pm$ 0.5	99.2 $\pm$ 0.8	98.3 $\pm$ 1.7	99.3 $\pm$ 0.7
Running Track	30	157	99.1 $\pm$ 0.4	98.8 $\pm$ 0.6	97.7 $\pm$ 1.0	99.2 $\pm$ 0.7
OA	–	–	91.5 $\pm$ 1.1	89.8 $\pm$ 1.0	86.8 $\pm$ 1.1	83.9 $\pm$ 1.1
AA	–	–	91.5 $\pm$ 3.1	89.8 $\pm$ 3.3	86.7 $\pm$ 3.8	84.0 $\pm$ 3.4



(a) Classification map generated using CKLADA (Proposed) with 20 training samples per class.



(b) Classification map generated using CKLFDA [2] with 20 training samples per class.

Grass-healthy	Grass-stressed	Grass-synthetic	Tree	Soil	Water	Residential	Commercial	Road	Highway	Railway	Parking Lot 1	Parking Lot 2	Tennis Court	Running Track
---------------	----------------	-----------------	------	------	-------	-------------	------------	------	---------	---------	---------------	---------------	--------------	---------------

Fig. 3: Classification maps obtained from the multi-source (hyperspectral and pseudo-waveform LiDAR) dataset using (a) the proposed CKLADA embedding and (b) a CKLFDA (bottom) embedding, both with an SRC classifier following the embedding, and with 20 samples per class for training.